LBNL-48641
UCB-PTH-01/27

Wilson Lines and Symmetry Breaking on Orbifolds

Lawrence J. Hall, Hitoshi Murayama and Yasunori Nomura

Department of Physics
University of California, Berkeley, California 94720
and
Theoretical Physics Group
Ernest Orlando Lawrence Berkeley National Laboratory, MS 50A-5101
University of California, Berkeley, California 94720

Abstract

A five-dimensional $SU(5)$ grand unified theory is constructed on the orbifold S^1/Z_2, with gauge symmetry breaking accomplished by a background gauge field. The theory is shown to be equivalent to one with gauge symmetry breaking by orbifold boundary conditions — the two pictures are related by a non-periodic gauge transformation. This effective field theory possesses localized explicit breaking of $SU(5)$ on one of the orbifold fixed points. Split multiplets on this fixed point are shown not to induce violations of unitarity in scattering amplitudes.

1 Introduction

Symmetry breaking has been one of the most important themes in particle physics, field theory, model building, and also in many other areas of physics. In most models of symmetry breaking in field theory, there is a scalar field, with a non-trivial quantum number under the symmetry, acquiring a vacuum expectation value, which is the order parameter. In the case that the symmetry is local, *i.e.* a gauge symmetry, the Higgs mechanism operates and the longitudinal components of the gauge fields are generated by "eating" the fields responsible for symmetry breaking. As is well-known, an explicit breaking of gauge symmetry, such as adding mass terms to gauge bosons, leads to an inconsistent theory, breaking unitarity at high energies. However, a spontaneous symmetry breaking does not modify the high-energy behavior of the theory and the theory remains consistent despite massive gauge bosons [1].

Recently, there has been a strong interest in studying symmetry breaking phenomena using symmetry-violating boundary conditions in extra dimensions [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. Grand-unified symmetry and supersymmetry can be broken by gauge non-invariant and non-supersymmetric boundary conditions, respectively, via the Scherk–Schwarz mechanism [18]. Natural questions are: what is the order parameter? What degrees of freedom are "eaten" by the gauge field? Is the breaking spontaneous or explicit? In the case of supersymmetry breaking, Marti and Pomarol suggested that the order parameter in the Scherk–Schwarz mechanism is the F-component of the radion field, the fluctuation of the size of the extra dimensions, while its fermionic component supplies the longitudinal components for the gravitino [19]. The questions we ask here are concerned with gauge symmetry breaking.

On multiply-connected manifolds, it is well-known that gauge symmetries may be broken by the presence of background gauge fields. This phenomenon has been known as the Hosotani mechanism, or Wilson-line symmetry breaking [20, 21]. We review this mechanism in section 2, as we wish to explore its relation to symmetry breaking by boundary conditions. This relationship is also discussed in [15]. When one classifies possible vacua for gauge fields on multiply-connected compact manifolds M, one finds that it is equivalent to the classification of flat bundles, which in turn can be classified by the "representations" (homomorphisms) of the fundamental group $\pi_1(M)$ into the gauge group. The associated gauge group element for each cycle is the holonomy. In section 3 we discuss how this idea naturally extends to orbifolds, where the concept of the fundamental group is ill-defined. Many of these concepts were introduced in the context of studying low energy limits of string theory [22]. In section 4 we discuss how discrete flavor orbifold symmetries of an effective field theory may arise from a more fundamental gauge theory at higher energies.

In section 5 we discuss the $SU(5)$ grand unified theory introduced in [8] where the origin of gauge symmetry breaking is viewed as a boundary condition imposed on fields on the orbifold S^1/Z_2. This theory was further studied in [10], where it was viewed as a theory with a restricted set of unbroken gauge symmetries in five dimensions, which did not include all the usual four-dimensional $SU(5)$ transformations. Here we obtain the identical theory with a non-zero background gauge field breaking the gauge symmetry rather than an orbifold boundary condition. We demonstrate that these two viewpoints are related by a hypercharge gauge transformation which is not periodic. While the background gauge viewpoint suggests that the gauge symmetry is broken spontaneously, in section 6 we stress that certain special explicit breaking of the gauge symmetry may be present. Non-unified multiplets and their interactions, respecting only the unbroken symmetries on the brane, are allowed, while the physics in the bulk is still constrained by the full gauge invariance. We check that this explicit breaking does not lead to additional violations of unitarity in certain scattering amplitudes. In section 7 we present our conclusions.

2 Holonomies on Manifolds

Let us first review Wilson Lines on compact manifolds. Wilson lines have been used to break gauge symmetries when extra dimensions are compactified [21, 23, 24].

Consider a gauge theory on a compact manifold M with the standard Hamiltonian

$$H = \int dx \frac{1}{2}\left[(\vec{E}^a)^2 + (\vec{B}^a)^2\right]. \tag{1}$$

The question is what the ground-state configurations are for the gauge field.

In the temporal (or Weyl) gauge where $A_0^a = 0$, the Hamiltonian simplifies to

$$H = \int dx \frac{1}{2}\left[(\dot{\vec{A}}^a)^2 + (\vec{B}^a)^2\right]. \tag{2}$$

For time-independent configurations, the first term simply vanishes. Because the Hamiltonian is positive semi-definite, the ground-state configuration is given by the equation

$$\vec{B}^a = 0. \tag{3}$$

On a simply-connected space (a connected space with no non-contractible cycles), the solution to this equation is simply that the gauge field is a pure-gauge configuration.

However, the situation is much more interesting on non simply-connected manifolds. Let us take a plane with the origin removed $M = \mathbb{R}^2\backslash\{0\}$ and $U(1)$ gauge group as an example. The condition that the field strength vanishes $B = 0$ allows the vector potential

$$(A_x, A_y) = \frac{\Phi}{2\pi}\vec{\nabla} \arctan\frac{y}{x} = \frac{\Phi}{2\pi}(-\frac{y}{r^2}, \frac{x}{r^2}). \tag{4}$$

Because the vector potential is a gradient, *i.e.* pure gauge, the field strength vanishes identically. There is, however, the Aharonov–Bohm phase

$$\exp\left(i\frac{e}{\hbar c}\oint \vec{A}\cdot d\vec{x}\right) = \exp\left(i\frac{e\Phi}{\hbar c}\right). \tag{5}$$

Unless $\Phi \neq 2\pi n\hbar c/e$ for $n \in \mathbb{Z}$, the gauge field has a physical significance. Unlike the Aharonov–Bohm effect, however, where one can imagine looking at the origin and measure the magnetic flux Φ there, the origin is *outside* the manifold in this example, and the magnetic flux is ill-defined. In fact, any magnetic fluxes which differ by an integer multiple of $2\pi\hbar c/e$ give identical physics on M. This is because one can perform a gauge transformation by

$$U = \exp\left(in \arctan\frac{y}{x}\right), \tag{6}$$

which is single-valued on M. We find

$$\vec{A} \to \vec{A} - i\frac{\hbar c}{e}U^{-1}\vec{\nabla}U = \vec{A} + n\frac{\hbar c}{e}(-\frac{y}{r^2}, \frac{x}{r^2}). \tag{7}$$

Therefore, Φ and $\Phi + 2\pi n\hbar c/e$ are gauge-equivalent. From this point on, we set $\hbar = c = 1$ and the coupling constant e absorbed in the definition of the gauge field.

It is sometimes more convenient to try to "gauge away" the gauge field as much as possible. For the above example, we can try

$$U = \exp\left(i\frac{\Phi}{2\pi}\arctan\frac{y}{x}\right). \tag{8}$$

The problem with this gauge transformation is that it is not single-valued. If we start from the positive x-axis and go around the origin counter-clockwise, the phase of U in Eq. (8) monotonically increases, becoming Φ when we come back to the positive x-axis from below. This is, in general, not an integer multiple of 2π and hence U is discontinuous. We need to put a "cut" along the positive x-axis. The gauge field is now identically zero all around, but when one crosses the cut, we need a gauge transformation by $\exp(i\Phi)$ ("transition function"). One can view this as a boundary condition for fields when going around the origin and do not have to refer to gauge field background anymore. Therefore, ground-state gauge field configurations can be classified by assigning Aharonov–Bohm phases or transition functions for non-contractible cycles without writing down explicit forms of the gauge fields.

Let us discuss a simple example of $M = S^1$ parameterized by $y \in [0, 2\pi R]$ and $G = SU(2)$. We take the gauge field to be along the 3rd isospin direction

$$A(y) = A^3(y)\frac{1}{2}\tau^3 = A^3(y)\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{9}$$

where A is the spatial component of the gauge field. Without loss of generality, one can make A^3 constant by use of gauge transformations. Moreover, the constant A^3 can be further gauge-transformed using the single-valued gauge transformation function

$$U(y) = \exp\left(i\tau^3\frac{y}{R}\right), \tag{10}$$

as

$$A^3 \to A^3 + \frac{2}{R}. \tag{11}$$

Therefore A^3R is defined mod 2. The holonomy, T, is given by

$$T = \exp\left(i\oint A^3\frac{\tau^3}{2}dy\right) = \begin{pmatrix} e^{\pi i A^3 R} & \\ & e^{-\pi i A^3 R} \end{pmatrix}. \tag{12}$$

The spectrum of particles is affected by the presence of the constant gauge field, or equivalently, a non-trivial transition function. If we use the language of the constant gauge field, the Kaluza–Klein (KK) modes of a Klein–Gordon field ϕ of isospin $I_3 = 1/2$ reads as

$$-\left(\partial_y - i\frac{1}{2}A^3\tau^3\right)^2\phi = m_{KK}^2\phi. \tag{13}$$

ϕ is periodic: $\phi(y+2\pi R) = \phi(y)$. Then the n-th mode has the form $\phi_n(y) = e^{iny/R}$, and

$$-\left(\partial_y - i\frac{1}{2}A^3\tau^3\right)^2\phi_n = \left(\frac{n}{R} - \frac{1}{2}A^3\tau^3\right)^2\phi_n. \tag{14}$$

Therefore the KK mass is $m_n = \left|\frac{n}{R} - \frac{1}{2}A^3\tau^3\right|$. The gauge transformation Eq. (10) would lead to a relabeling of states by $n \to n \pm 1$, but the physical spectrum remains unchanged as it should. If instead we had gauged away the gauge field by

$$U(y) = \exp\left(-\frac{i}{2}A^3\tau^3 y\right), \tag{15}$$

the mode equation is simply

$$-\partial_y^2\phi = m_{KK}^2\phi, \tag{16}$$

while the boundary condition is changed to

$$\phi(2\pi R) = e^{-\pi i A^3\tau^3 R}\phi(0). \tag{17}$$

The modes are given by

$$\phi_n(y) = \exp\left[i\left(\frac{n}{R} - \frac{1}{2}A^3\tau^3\right)y\right], \tag{18}$$

giving the same spectra as obtained using a constant gauge field and periodic ϕ.

The KK spectrum of the gauge field is also affected by the holonomy. Because A^3_μ commutes with A^3 itself, the spectrum of the unbroken $U(1)$ is simple. Here, μ represents the index of uncompactified spacetime dimensions. The mode functions are $A^3_{\mu,n}(y) = \cos ny/R$ and $\sin ny/R$, and the mass spectrum is n/R with a single zero mode: the gauge field of the unbroken $U(1)$. On the other hand, the $W^\pm$-bosons become massive. Their isospins are $I_3 = \pm 1$, and the same analysis for the Klein–Gordon field leads to the KK W-spectrum of $m_n = \left|\frac{n}{R} \mp A^3\right|$.

In general, vanishing field-strength configurations (or "flat connections") on non-simply connected manifolds can be classified in terms of Aharonov–Bohm phases ("holonomies") for each non-contractible cycles. "Magnetic fluxes" are "outside" the manifold and have no physical meaning; however the holonomies do. The cycles can be continuously deformed without changing the phase, because the difference in the phase can be expressed as an area integral of the magnetic field using Stokes' theorem which vanishes because of the vanishing field strength. Therefore a holonomy is associated with a class of non-contractible cycles that can be continuously deformed among each other. For each class of non-contractible cycles c_i, we associate a gauge group element $g_i \in G$ by

$$g_i = \mathcal{P}\exp\left(i\oint_{y_0,c_i} A(y)dy\right), \tag{19}$$

where $\mathcal{P}$ denotes the path-ordered product. Note that the cycles are defined with respect to an origin y_0 (even though the group itself does not depend on the choice of the origin) and the associated holonomies are also defined with respect to the origin. Under gauge transformations, the elements transform as

$$g_i \to U(y_0)g_iU(y_0)^{-1}. \tag{20}$$

Just like in the abelian example at the beginning of the section, g_i does not change under the continuous deformation of the cycle c_i because the field strength vanishes.

When one goes around non-contractible cycles successively, in some cases it matters in which order one goes around different cycles. Successive loops can be viewed as a product of two cycles, defining a group among non-contractible cycles: the fundamental group or the first homotopy group $\pi_1(M)$. In the above example of a plane with the origin removed, $\pi_1(M) = \mathbb{Z}$ because a non-contractible cycle is just labeled by the number of times it goes around the origin. For more complicated manifolds, such as a genus-two Riemann surface (two-hole doughnut), the fundamental group is non-abelian.

The ground-state configurations of the gauge field are classified by the gauge-inequivalent classes of holonomies. In other words,

$$\{\text{inequivalent ground state configurations}\} = \text{Hom}(\pi_1(M) \to G)/G, \tag{21}$$

where the r.h.s. is the set of all homomorphisms (maps that respect multiplications of elements) from the fundamental group $\pi_1(M)$ to the gauge group G, modded out by the gauge equivalence G.

When the manifold M is defined as a coset M_0/H of a simply-connected manifold M_0 by a discrete symmetry group H, the fundamental group is simply H itself. This is the consequence of the homotopy exact sequence

$$\cdots \to \pi_n(M_0) \to \pi_n(M_0/H) \to \pi_{n-1}(H) \to \pi_{n-1}(M_0) \to \cdots. \tag{22}$$

Because of the assumption that M_0 is simply-connected, $\pi_1(M_0) = \pi_0(M_0) = 0$,[1] we obtain the exact sequence

$$0 = \pi_1(M_0) \to \pi_1(M_0/H) \to \pi_0(H) \to \pi_0(M_0) = 0. \tag{23}$$

For this to be an exact sequence, we find

$$\pi_1(M_0/H) = \pi_0(H) = H, \tag{24}$$

where the last equality holds because we assumed that H is a discrete group. Then the space of ground-state configurations is

$$\{\text{inequivalent ground state configurations}\} = \text{Hom}(H \to G)/G, \tag{25}$$

or equivalently, a set of all possible "representations" of the discrete group H in G modulo gauge equivalences.

It is instructive to look at an example where H is a discrete group. For instance, consider $M_0 = S^2$, and $H = Z_2$ identifies the anti-podal points of the sphere,

[1] The zeroth homotopy group π_0 refers to the disconnected parts of the space.

so that $M = S^2/Z_2 = \mathbb{R}P^2$.[2] On S^2, scalar fields (or gauge fields in remaining uncompactified dimensions) are expanded in terms of spherical harmonics Y_l^m with $m^2 = l(l+1)$. Note that under Z_2, $Y_l^m \to (-1)^l Y_l^m$. On $\mathbb{R}P^2$, one can have the holonomy $T = \mathrm{diag}(-1,-1,+1)$ in the $SU(3)$ gauge group. The main difference from the S^1 case discussed earlier in this section is that the choice for T is discrete. Then the unbroken $SU(2)\times U(1)$ gauge bosons have even T, and hence only even l are allowed, while the broken gauge bosons have odd T, and only odd l are allowed. This does not allow the broken gauge bosons to be massless as expected.

3 Holonomies on Orbifolds

An orbifold is defined as a coset of a manifold by a discrete symmetry. If the symmetry does not act freely, *i.e.* there are fixed points that do not move under the symmetry, the coset space is singular at the fixed points and is not a manifold. It is an orbifold instead. Because it is not a manifold, the conventional definition of the homotopy groups fails. Therefore, we cannot classify the ground-state configurations of the gauge field using Eq. (21). However, Eq. (25) is still well-defined because the very definition of an orbifold is in terms of a coset space.

For example, consider the $S^1/\mathbb{Z}_2$ orbifold. This is a circle $y \in [0, 2\pi R]$ modded out by the "parity" $\mathbb{Z}_2$ symmetry $P: y \to -y$. Of course $P^2 = 1$. We cannot apply Eq. (25) because S^1 is not simply connected. However, S^1 itself can be viewed as a coset space $\mathbb{R}/\mathbb{Z}$, where $\mathbb{R}$ is the real axis for y and $\mathbb{Z}$ is generated by the translation $T: y \to y + 2\pi R$. Then the orbifold $S^1/\mathbb{Z}_2 = \mathbb{R}/(\mathbb{Z} \ltimes \mathbb{Z}_2)$. The translation T and the parity do not commute,

$$PTP = T^{-1}, \tag{26}$$

and hence the semi-direct product $H = \mathbb{Z} \ltimes \mathbb{Z}_2$. Because $\mathbb{R}$ is simply-connected, we can now use Eq. (25) to find the space of ground-state configurations: all representations of $\mathbb{Z} \ltimes \mathbb{Z}_2$ in the gauge group G. Note that Eq. (26), together with $P^2 = 1$, defines the dihedral group D_n if $T^n = 1$. Therefore the orbifold group here is the $n \to \infty$ limit of the dihedral groups. The holonomy T in this example can be represented as a Wilson line in exactly the same fashion as on manifolds,

$$T = \mathcal{P} \exp\left(i \oint_0^{2\pi R} A(y) dy \right), \tag{27}$$

where $\mathcal{P}$ denotes the path-ordered product. However, the unitary matrix representing P cannot be expressed in terms of a gauge-field background.

[2] This manifold is unorientable, but it does not matter for the purpose of this discussion.

In general, elements for freely acting generators of H can be expressed in terms of background gauge fields, while those for non-freely acting generators of H (those with fixed points) cannot be. Note that, one can view the whole theory to be G invariant, with the non-trivial holonomy as the "expectation values" of the gauge fields. Even though the holonomy for non-freely acting generators of H cannot be expressed in terms of the background gauge fields, they are a part of the classification of the symmetry breaking as given in Eq. (25).

When $G = U(1)$, both T and P are represented as phases, and hence they commute. Even when G is non-abelian, T and P may commute for certain representations and the following discussions still apply. Then $T^2 = 1$ from Eq. (26) and only possibilities are $T = \pm 1$. Because $P^2 = 1$, we also find $P = \pm 1$. Let us check this result based on a formal argument explicitly using $G = SU(2)$ as an example. We take the gauge field to be along the 3rd isospin direction as in the previous section,

$$A(y) = A^3(y)\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{28}$$

Again without loss of generality, one can make A^3 constant by use of gauge transformations. Moreover, the constant A^3 can be further gauge-transformed using Eq. (10) as

$$A^3 \to A^3 + \frac{2}{R}. \tag{29}$$

Therefore A^3R is defined mod 2. Under the parity, the gauge field $A(y)$ transforms as $A(y) \to -A(-y)$, because the covariant derivative $\partial_y - iA(y)$ must transform covariantly. (Here and hereafter, we normalize the gauge field such that the kinetic term has the coefficient $1/g^2$ while the covariant derivative does not have the coupling constant.) This fixes P on the gauge field uniquely to be $P = -1$. Naively, it appears that the gauge field must vanish identically because of the parity invariance. However, it is possible that the gauge field transforms to its gauge-equivalent configuration under parity, and hence

$$A^3 = \frac{1}{R}, \tag{30}$$

is perfectly allowed. Under parity $y \to -y$ it transforms to $-1/R$, but is gauge equivalent to $+1/R$ by the use of the gauge transformation Eq. (10). In other words, if we redefine the parity as the combined transformation of $y \to -y$ and the gauge transformation Eq. (10), the vacuum configuration Eq. (30) is invariant. Then the holonomy for going around the circle is

$$T = \exp\left(i \oint A^3 \frac{\tau^3}{2} dy\right) = -1. \tag{31}$$

Therefore Eq. (30) is a consistent ground-state configuration of the gauge field which corresponds to the possibility found from the formal argument. It is interesting to note that this choice of T is in the center of the gauge group, and hence does not break the gauge invariance. Another way to see it is that $T = -1$ on the fundamental representation (doublet) is trivial, $T = 1$, on integer isospin representations. The W-bosons with the spectrum $m_n = \left|\frac{n}{R} \mp A^3\right|$ indeed have zero modes. The only effect then is on half-odd isospin fields that can be viewed simply as anti-periodic boundary conditions for them under the shift T.

The fact that the gauge group is not broken with Eq. (30) is a special situation with the $SU(2)$ gauge group where the only non-trivial choice for the holonomy was in the center. To break the $SU(2)$ gauge invariance, we have to consider the case with $A^3 = 1/(2R)$. This is possible if the theory possesses the discrete gauge symmetry under which all the half-odd isospin fields ϕ transform as $\phi \to \pm i\phi$. Then, combined with the non-trivial holonomy in the flavor space, $\phi \to \pm i\phi$, we can break the $SU(2)$ gauge symmetry. A similar situation arises when one tries to achieve, *e.g.*, $SU(2n) \to SU(n) \times SU(n) \times U(1)$ breaking. This issue is discussed in the next section.

An example which breaks the gauge invariance by Eq. (30) is $SU(5)$ with $T = \mathrm{diag}(1, 1, 1, -1, -1)$, which has interesting phenomenological applications. We will see this example in more detail in section 5. A somewhat simpler example is the $SU(3)$ gauge group with the holonomy $T = \mathrm{diag}(-1, -1, 1)$. With the choice analogous to that in Eq. (30), the gauge group is indeed broken to $SU(2) \times U(1)$, and it is easy to work out the spectrum and verify explicitly that the gauge bosons for the broken generators indeed do not have zero modes.

Another interesting possibility is a non-commuting P and T. For instance, when $G = SO(3)$ (or $O(2)$), the following representation is possible, motivated by the connection to the dihedral groups:

$$P = \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \qquad T = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{32}$$

Such a possibility would be useful to reduce the rank of the gauge group.

4 Flavor Holonomy

For matter fields, there are often additional flavor symmetries. Then one can assign non-trivial holonomies in the flavor space as well. For instance, a real scalar field with the potential $V = m^2\phi^2 + \lambda\phi^4$ has a $\mathbb{Z}_2$ symmetry $\phi \to -\phi$. Then for each non-contractible cycle, we can assign either even or odd $\mathbb{Z}_2$ element under the flavor symmetry, *i.e.* periodic or anti-periodic boundary conditions.

When can we assign holonomies for non-gauge symmetries? We can do so when we can regard the flavor symmetry "gauged" in the following sense [25]. The above $\mathbb{Z}_2$ model can be promoted to a full $U(1)$ model by adding an imaginary part to the scalar field $\phi + i\eta$. The $U(1)$ gauge group is broken to $\mathbb{Z}_2$ by an expectation value of the field ψ with charge $+2$. Without loss of generality, we can take $\langle\psi\rangle \in \mathbb{R}$. Then using a term $(\phi - i\eta)^2\langle\psi\rangle + (\phi + i\eta)^2\langle\psi^*\rangle = 2(\phi^2 - \eta^2)\langle\psi\rangle$, we can make ϕ and η non-degenerate; the low-energy effective theory may well have only ϕ but not η. On the other hand, the $U(1)$ gauge group is broken by the charge two field and hence allows a "cosmic string" solution where the ψ field has its phase rotated by 2π as one goes around a non-contractible cycle. For a charge one field, however, the phase rotates by only π; this situation is equivalent to having the holonomy of -1 under the $U(1)$ gauge group. In this sense, the anti-periodic boundary condition can be viewed as a non-trivial holonomy of the "gauged $\mathbb{Z}_2$ symmetry". Once taken this view, there are non-trivial constraints on what flavor holonomies can exist. For instance, the flavor symmetry must be anomaly-free [26].

In some cases, the existence of the non-trivial holonomy of gauged discrete symmetries is essential to break gauge symmetries. (This corresponds to the case $P_G^2 = P_H^2 = -1$ in [14].) This happens, for instance, in the case $G = SU(2)$. As we discussed in the previous section, if there were no discrete gauge symmetry, the gauge transformation function which does not change the boundary conditions of the fields is given by Eq. (10). Then, the only vacuum configuration allowed would be Eq. (30), since the gauge transformation of A^3 is given by Eq. (29). However, if the $SU(2)$ gauge theory possesses the $\mathbb{Z}_4$ discrete gauge symmetry under which all the half-odd isospin fields have charges ± 1, then we can use the non-trivial holonomy of this $\mathbb{Z}_4$ symmetry to change the situation. Specifically, we can consider the vacuum configuration

$$A^3 = \frac{1}{2R}, \tag{33}$$

together with the boundary condition for half-odd isospin fields ϕ

$$\phi(y + 2\pi R) = \pm i\phi(y), \tag{34}$$

since these vacuum configuration and boundary conditions are invariant under the "parity" that is defined as the combined transformation of $y \to -y$ and

$$U(y) = \exp\left(i\tau^3 \frac{y}{2R}\right). \tag{35}$$

In this case, the holonomy in the gauge space is given by

$$T_G = \exp\left(i \oint A^3 \frac{\tau^3}{2} dy\right) = i\tau^3, \tag{36}$$

and thus the $SU(2)$ gauge symmetry is broken. On the other hand, the holonomy in the flavor space is

$$T_F = \pm i, \tag{37}$$

for the half-odd isospin fields ($T_F = \pm 1$ for integer isospin fields). Therefore, the consistency condition $T^2 = (T_G \otimes T_F)^2 = 1$ still holds (all the fields transform as ± 1 under T).

5 $SU(5)$ Breaking on S^1/Z_2

In this section we reproduce the $SU(5)$ gauge breaking introduced in [8] using the language of the vacuum configuration. The model is based on a five-dimensional theory with the gauge group $SU(5)$ broken by non-trivial orbifold boundary conditions. The structure and phenomenology of this type of theory were discussed in detail in [10], and several models have been constructed so far [8, 9, 10, 12, 13].

The $SU(5)$ gauge transformation acts on various fields as

$$\phi(x^\mu, y) \quad \to \quad U\phi(x^\mu, y), \tag{38}$$

$$A_5^a T^a(x^\mu, y) \quad \to \quad U A_5^a T^a(x^\mu, y) U^{-1} - iU^{-1}\partial_y U, \tag{39}$$

where $U = \exp(i\theta^a(x, y)T^a)$ and A_5 is the 5th component of the gauge field. To achieve the desired breaking of the $SU(5)$ symmetry, we consider the vacuum configuration such that A_5 has an expectation value in the hypercharge direction. Therefore we concentrate on this direction from now on. Then, the above transformations can be written as

$$\phi(x^\mu, y) \quad \to \quad \mathrm{e}^{i\theta(x^\mu, y)Y_\phi}\phi(x^\mu, y), \tag{40}$$

$$A_5(x^\mu, y) \quad \to \quad A_5(x^\mu, y) + \partial_y\theta(x^\mu, y). \tag{41}$$

Here, A_5 and θ are in the hypercharge direction and the hypercharges Y_ϕ of ϕ are normalized such that they take half integer values.

As in section 3, we consider the large gauge transformation with $\theta(x^\mu, y)$ linear in y. Then, from the periodicity of the gauge transformation under $y \to y + 2\pi R$,

$$\theta(x^\mu, y) = \frac{2n}{R} y, \tag{42}$$

where $n \in \mathbb{Z}$. Thus, the gauge transformation on A_5 is given by

$$A_5(x^\mu, y) \to A_5(x^\mu, y) + \frac{2n}{R}, \tag{43}$$

so that the physically inequivalent vacuum configuration is parameterized as

$$A_5 \equiv \langle A_5(x^\mu, y)\rangle = -\frac{a}{R}, \tag{44}$$

where $0 \leq a < 2$.

Under the orbifold parity $y \to -y$, A_5 has an odd transformation property. This quantizes the possible values of a to be 0 or 1. The $a = 0$ case is obviously allowed. In the case of $a = 1$, A_5 changes the sign under $y \to -y$, but the transformed value is gauge equivalent to the original value, so that this case is also allowed. In other words, we can redefine the "parity" as the combined transformation of $y \to -y$ and Eq. (42) such that the vacuum configuration is invariant under this "parity". The breaking of $SU(5)$ only occurs when $a = 1$. Thus we concentrate on the case $a = 1$ from now. This gives the holonomy

$$T = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & -1 & \\ & & & & -1 \end{pmatrix}, \tag{45}$$

which clearly breaks $SU(5)$ down to $SU(3) \times SU(2) \times U(1)$.

Let us now work out the wavefunctions for various fields. Under the vacuum configuration Eq. (45), the "parity" is defined by the combined transformation of $y \to -y$ and the large gauge transformation given by Eq. (42). Therefore the boundary conditions for the fields ϕ_+ (ϕ_-) that transformed as even (odd) under the $y \to -y$, such as A_μ (A_5), are now given by

$$\phi_\pm(x^\mu, -y) = \pm e^{i\frac{2Y_\phi}{R}y}\phi_\pm(x^\mu, y). \tag{46}$$

The mode expansions which satisfy the above boundary conditions are easily found as

$$\phi_+(x^\mu, y) = \sum_{n=0}^{\infty} \phi_{+n}(x^\mu)\, e^{-i\frac{Y_\phi}{R}y} \cos\frac{n}{R}y \qquad \text{for } Y_\phi : \text{integer}, \tag{47}$$

$$\phi_+(x^\mu, y) = \sum_{n=0}^{\infty} \phi_{+n}(x^\mu)\, e^{-i\frac{Y_\phi}{R}y} \cos\frac{n+\frac{1}{2}}{R}y \qquad \text{for } Y_\phi : \text{half integer}, \tag{48}$$

$$\phi_-(x^\mu, y) = \sum_{n=0}^{\infty} \phi_{-n}(x^\mu)\, e^{-i\frac{Y_\phi}{R}y} \sin\frac{n+1}{R}y \qquad \text{for } Y_\phi : \text{integer}, \tag{49}$$

$$\phi_-(x^\mu, y) = \sum_{n=0}^{\infty} \phi_{-n}(x^\mu)\, e^{-i\frac{Y_\phi}{R}y} \sin\frac{n+\frac{1}{2}}{R}y \qquad \text{for } Y_\phi : \text{half integer}. \tag{50}$$

Substituting these expansions into the original five-dimensional action, we obtain the KK mass spectrum $m_n = n/R$, $m_n = (n+1/2)/R$, $m_n = (n+1)/R$ and $m_n = (n+1/2)/R$ for each of the above four cases. Note that the phase factor ensures that all modes are periodic under $y \to y + 2\pi R$.

The KK spectrum of the gauge fields is easily identified. For the unbroken gauge generators of $SU(3)\times SU(2)\times U(1)$, the gauge fields do not have hypercharge so that the spectrum is given by $m_n = n/R$. On the other hand, for the broken generators (X, Y bosons), their hypercharge is 5/6 (and hence half integer in the normalization here), and the mass spectrum is $m_n = (n+1/2)/R$. This therefore gives exactly the same mass spectrum worked out in the original paper [8] using the language of orbifold boundary conditions.

One of the attractive features of this type of model is that the doublet-triplet splitting can be achieved quite nicely if the Higgs fields live in the bulk. One might naively think that since the triplet (doublet) Higgses have integer (half integer) hypercharges, the triplet Higgses have zero modes but the doublet ones do not. However, this situation can easily be changed by introducing a discrete $\mathbb{Z}_2$ gauge symmetry acting on the Higgs fields (see also [14]). Then, we can have a flavor holonomy in addition to the gauge holonomy Eq. (45), and the wavefunctions for the integer and half integer hypercharge fields are interchanged in Eqs. (47 – 50). Therefore, we have massless Higgs doublets in this case. From a single Higgs hypermultiplet in **5** representation, we have a freedom to assign even parity to **5** and odd parity to $\mathbf{5}^*$ chiral multiplets or vice versa. For the first choice, we find the mass spectrum

$$\begin{array}{c|c|c|c} H_D & H_D^c & H_T & H_T^c \\ \hline n/R & (n+1)/R & (n+\frac{1}{2})/R & (n+\frac{1}{2})/R \end{array} \tag{51}$$

where H_D is identified with the up-type Higgs field. The opposite parity assignment gives the down-type Higgs field as a massless mode. This complete the derivation of KK mass spectrum in the orbifold theory in terms of the language of vacuum configuration.

Finally, we discuss the relation between the wavefunctions given in Eqs. (47 – 50) and those given in the orbifold boundary condition picture [8, 5], which does not involve any phase factor. The above wavefunctions are derived on the constant background of non-zero A_5. However, we can gauge away this background by making a gauge transformation

$$\theta(x^\mu, y) = \frac{y}{R}. \tag{52}$$

In this gauge, the phase factors in Eqs. (47 – 50) are removed and the wavefunctions in the orbifold boundary condition picture are reproduced. This gauge transformation is not periodic; hence, in the picture with $A_5 = 0$, the fields have non-trivial boundary conditions under $y \to y + 2\pi R$. The holonomy Eq. (45) is transformed from a property of the background gauge field to a boundary condition. The two pictures describe the same theory, and are related by a non-periodic gauge transformation.

6 Symmetry Breaking on Orbifold Fixed Points

In this section we discuss several issues in the orbifold breaking of gauge symmetries. As an example, we consider the models of $SU(5)$ breaking [8, 9, 10, 12, 13] where there is a non-trivial holonomy for the translation T but not for Z. In this case, the relation between the two pictures of Wilson-line and orbifold symmetry breakings are particularly simple: the orbifold breaking picture corresponds to taking a special gauge where the constant vacuum expectation value of A_5 in the bulk is kept equal to zero and the effects of the non-trivial holonomy are all represented by boundary conditions.

In the setup of orbifold symmetry breaking, there exists a point where only a subgroup of G is manifest, and on that point incomplete multiplets and/or G non-invariant interactions can be introduced [10]. It has recently been proposed to solve the doublet-triplet splitting problem by putting just the Higgs doublets on the fixed point $y = \pi R$ where only $SU(3) \times SU(2) \times U(1)$ is manifest [12]. One should wonder if such an "explicit" breaking of the $SU(5)$ gauge invariance would lead to any inconsistencies. However, below we show that such incomplete multiplets pass non-trivial consistency checks, as required in gauge theories, in both pictures.

First, we emphasize that these explicit breakings are allowed only on the fixed point where only a subgroup of $SU(5)$ is manifest. The bulk must still be $SU(5)$ symmetric to keep unitarity up to the cutoff of the theory. For instance, the gauge coupling constants in the bulk must be unified into $SU(5)$. Second, in order for the theory to make sense without color-triplet Higgs on the brane, there should not be any transitions between the doublet and color-triplet Higgses. Indeed, in both pictures the X and Y bosons have wavefunctions proportional to $\cos[(n+1/2)y/R]$, and hence they vanish on the brane and cannot cause transitions. Note that in the example of a manifold with a discrete holonomy (*e.g.*, $\mathbb{R}P^2$), there is no point where the wavefunction of all broken gauge bosons (Y_l^m with odd l) identically vanish. This possibility is special to orbifolds.

Third, the partial wave unitarity is satisfied up to the cutoff of the theory. Consider the process $H_D H_D^* \to X X^*$ where X is the X-boson. In four-dimensional field theory, this process can proceed through the s-channel exchange of $U(1)_Y$ or $SU(2)_L$ gauge bosons, and also through the t-channel color-triplet Higgs exchange. Each diagram behaves as $g^2 E^2/m_X^2$ at high energies $E \gg m_X$ in $J = 1$ partial wave, which would cause violation of partial wave unitarity. However, two diagrams cancel and one obtains an amplitude of $\sim g^2 \log(E^2/m_X^2)$, which remains unitarity up to extremely high energies. In our orbifold setup, however, the color-triplet Higgs does not exist and hence only the first diagram exists. It appears that partial wave unitarity is violated at an energy scale $E \sim \sqrt{4\pi} m_X/g$. It turns out, however, that not only the zero mode but also the first mode of the KK-tower of $U(1)_Y$ and $SU(2)_L$ gauge bosons can be exchanged conserving KK-momentum at the triple-gauge-boson vertex. The triple-gauge-boson vertex for the zero mode is determined by the integral of wavefunctions

$$g_5 \int_0^{\pi R} dy \left(\sqrt{\frac{2}{\pi R}} \cos \frac{y}{2R} \right)^2 \sqrt{\frac{1}{\pi R}} = g_5 \sqrt{\frac{1}{\pi R}} = g_4, \tag{53}$$

while that for the first KK mode

$$g_5 \int_0^{\pi R} dy \left(\sqrt{\frac{2}{\pi R}} \cos \frac{y}{2R} \right)^2 \sqrt{\frac{2}{\pi R}} \cos \frac{y}{R} = g_5 \sqrt{\frac{1}{2\pi R}} = \frac{g_4}{\sqrt{2}}. \tag{54}$$

On the other hand, the gauge coupling of the zero mode to the Higgs doublet on the brane is

$$g_5 \sqrt{\frac{1}{\pi R}} \Bigg|_{y=\pi R} = g_4, \tag{55}$$

while that of the first KK mode is

$$g_5 \sqrt{\frac{2}{\pi R}} \cos \frac{y}{R} \Bigg|_{y=\pi R} = -\sqrt{2}\, g_4. \tag{56}$$

Therefore, the diagram with the first KK mode exchanged in the s-channel has the coupling factor $(g_4/\sqrt{2})(-\sqrt{2}\, g_4) = -g_4^2$ as opposed to g_4^2 in the diagram with the zero mode. The cancellation between the two diagrams leads to a high-energy behavior for the amplitude of $\sim g^2(E^2/m_X^2)(m_{KK}^2/E^2) \sim g^2$ which preserves unitarity at high energies. One can also check that amplitudes for any combination of KK X-bosons in the final state satisfy partial wave unitarity. Of course, the multiplicity ER of KK X-boson states lead to the factor ER in amplitude for the fastest-growing channel, and would violate unitarity at $g^2 ER \sim 4\pi$, or $E \sim 4\pi/(g^2 R)$, as expected in any five-dimensional gauge theories. The point is that the incomplete multiplet on the brane does not lead to any additional unitarity violation beyond what is expected in five-dimensional gauge theories.

One way of realizing a split multiplet on the brane is through a non-local operator using the Wilson lines. For example, in the supersymmetric $SU(5)$ theory, the gauge fields in the 5th dimension can be represented as a chiral superfield $\Sigma(y)$ in the adjoint representation, and hence one can write down a coupling on the brane

$$\int d^2\theta \bar{H} \left(M - \mu \mathcal{P} e^{\oint dy \Sigma(y)} \right) H, \tag{57}$$

where the Higgs multiplets, H and $\bar{H}$, fill complete $SU(5)$ multiplets of $\mathbf{5}$ and $\bar{\mathbf{5}}$ and are localized on the brane, while the Wilson line is the integral over the 5th dimension. By a fine-tuning between M and μ, we can achieve split multiplets on the brane. This operator can be used even if the brane does not sit at the fixed point. In this case, the unitarity consideration indeed leads to inconsistencies unless we include the effect of the color-triplet Higgs. One also sees the same unitarity violation if the brane is in S^1 rather than an orbifold. Therefore, the incomplete multiplet does not arise from this effective operator, but by putting it on the brane "by hand".

Note that we have discussed the incomplete multiplet of $SU(5)$ as an example, but the same argument applies for more general cases. In particular, we can introduce fields on the brane which do not come from a $SU(5)$ representation; we can even introduce fields whose $U(1)$ charges are not quantized.

Having passed many non-trivial consistency checks, we therefore say that explicit $SU(5)$ breaking is allowed on the brane while the bulk is still $SU(5)$ symmetric. Clearly, orbifolds allow wider possibilities beyond conventional spontaneous breaking of symmetries. No matter whether we use background gauge fields or orbifold boundary conditions, the X and Y gauge bosons $A_\mu^X(x,y)$ vanish at $y = \pi R$. This implies that the theory is only gauge invariant under a restricted set of gauge transformations $\theta^X(x,y)$. In particular $\theta^X(x,y)$ has no zero mode.

We finally comment on the relation between the $SU(5)$ breaking "vacuum" and the $SU(5)$ preserving "vacuum", *i.e.* $a = 1$ and $a = 0$ in Eq. (44). In the $SU(5)$ preserving "vacuum", we clearly cannot introduce incomplete multiplets on the brane, since it violates unitarity below the cutoff scale. Then, how do the split multiplets appears when we move from the $SU(5)$ preserving to the $SU(5)$ violating "vacua"? The answer in the effective field theory is that we simply cannot move between two "vacua". The potential barrier between two "vacua" is infinity, so that they are completely disconnected: there is no physical process which causes tunneling between the two. Therefore within the field theory there is no contradiction to have two disconnected "vacua" with different particle content; they are just two different theories. One may wonder if two "vacua" may be somehow connected in more fundamental theory. For instance, vacua with different number of generations can be connected in string theory. It does not cause a contradiction, however, because the singular point that connects two vacua with different number of generations gives a conformal theory with no particle interpretation [27].

7 Conclusions

We have constructed a $SU(5)$ grand unified theory on the orbifold S^1/Z_2 where all fields have $SU(5)$ conserving periodic boundary conditions, but the $SU(5)$ gauge symmetry is broken by a background gauge field, A_5. This is an example of the well-known Wilson line breaking mechanism. The Wilson line is non-trivial and is gauge invariant under periodic gauge transformations. However, if we perform a non-periodic gauge transformation we obtain a very different picture: the gauge field can be made to vanish while the symmetry breaking now appears in the orbifold boundary conditions. The physics of these two pictures is the same, since they contain the same spectra of KK towers and the same interactions.

An $SU(5)$ effective field theory on S^1/Z_2 can take two possible forms: $SU(5)$ may be broken or unbroken. In the unbroken case the Wilson line and orbifold boundary conditions both preserve $SU(5)$ in the same gauge, while in the broken case they do not. Working in the gauge with periodic boundary conditions, the question of $SU(5)$ breaking appears to rest on whether the gauge field A_5 has a vacuum expectation value or not — and it is tempting to say that these are two vacua of the same effective field theory. However, we have argued that the two effective field theories are different. One can have explicit $SU(5)$ breaking multiplets and interactions localized at an orbifold fixed point while the other cannot. For the broken case the zero mode gauge transformation $\theta^X(x)$ is not a symmetry of the theory, while in the unbroken case it is. The effective field theory for the broken case has the four-dimensional $SU(5)$ symmetry explicitly broken, and a restricted set of five-dimensional $SU(5)$ transformations which are exact and unbroken.

Even though the two effective field theories are different, it may well be that they are the low energy limit of different vacuum choices of a single more fundamental theory at higher energies. It could be that in this theory both the background gauge field and the symmetry breaking localized on the orbifold fixed point are generated spontaneously.

Acknowledgments

We thank Martin Schmaltz, David Smith and Neal Weiner for useful discussions. Y.N. thanks the Miller Institute for Basic Research in Science for financial support. This work was supported by the Department of Energy under contract DE-AC03-76SF00098 and the National Science Foundation under contract PHY-95-14797.

References

[1] C. H. Llewellyn Smith, Phys. Lett. B **46**, 233 (1973);
J. S. Bell, Nucl. Phys. B **60**, 427 (1973);
J. M. Cornwall, D. N. Levin and G. Tiktopoulos, Phys. Rev. Lett. **30**, 1268 (1973); Phys. Rev. D **10**, 1145 (1974) [Erratum-ibid. D **11**, 972 (1974)].

[2] I. Antoniadis, C. Munoz and M. Quiros, Nucl. Phys. B **397**, 515 (1993) [hep-ph/9211309].

[3] A. Pomarol and M. Quiros, Phys. Lett. B **438**, 255 (1998) [hep-ph/9806263];
A. Delgado, A. Pomarol and M. Quiros, Phys. Rev. D **60**, 095008 (1999) [hep-ph/9812489].

[4] I. Antoniadis, S. Dimopoulos, A. Pomarol and M. Quiros, Nucl. Phys. B **544**, 503 (1999) [hep-ph/9810410].

[5] R. Barbieri, L. J. Hall and Y. Nomura, Phys. Rev. D **63**, 105007 (2001) [hep-ph/0011311].

[6] N. Arkani-Hamed, L. Hall, Y. Nomura, D. Smith and N. Weiner, Nucl. Phys. B **605**, 81 (2001) [hep-ph/0102090].

[7] A. Delgado and M. Quiros, hep-ph/0103058.

[8] Y. Kawamura, hep-ph/0012125.

[9] G. Altarelli and F. Feruglio, Phys. Lett. B **511**, 257 (2001) [hep-ph/0102301].

[10] L. Hall and Y. Nomura, hep-ph/0103125.

[11] Y. Nomura, D. Smith and N. Weiner, hep-ph/0104041.

[12] A. Hebecker and J. March-Russell, hep-ph/0106166.

[13] R. Barbieri, L. J. Hall and Y. Nomura, hep-ph/0106190.

[14] R. Barbieri, L. J. Hall and Y. Nomura, hep-th/0107004.

[15] A. Hebecker and J. March-Russell, hep-ph/0107039.

[16] J. Bagger, F. Feruglio and F. Zwirner, hep-th/0107128.

[17] A. Masiero, C. A. Scrucca, M. Serone and L. Silvestrini, hep-ph/0107201.

[18] J. Scherk and J. H. Schwarz, Phys. Lett. B **82**, 60 (1979); Nucl. Phys. B **153**, 61 (1979).

[19] D. Marti and A. Pomarol, hep-th/0106256.

[20] Y. Hosotani, Phys. Lett. B **126**, 309 (1983); Phys. Lett. B **129**, 193 (1983).

[21] P. Candelas, G. T. Horowitz, A. Strominger and E. Witten, Nucl. Phys. B **258**, 46 (1985);
E. Witten, Nucl. Phys. B **258**, 75 (1985).

[22] L. E. Ibanez, H. P. Nilles and F. Quevedo, Phys. Lett. B **187**, 25 (1987); Phys. Lett. B **192**, 332 (1987).

[23] J. D. Breit, B. A. Ovrut and G. C. Segre, Phys. Lett. B **158**, 33 (1985);
A. Sen, Phys. Rev. Lett. **55**, 33 (1985);
M. Dine, V. Kaplunovsky, M. Mangano, C. Nappi and N. Seiberg, Nucl. Phys. B **259**, 549 (1985).

[24] J. P. Derendinger, L. E. Ibanez and H. P. Nilles, Nucl. Phys. B **267**, 365 (1986);
B. R. Greene, K. H. Kirklin, P. J. Miron and G. G. Ross, Nucl. Phys. B **278**, 667 (1986); Nucl. Phys. B **292**, 606 (1987);
S. Ferrara, C. Kounnas and M. Porrati, Phys. Lett. B **206**, 25 (1988);
A. Font, L. E. Ibanez, F. Quevedo and A. Sierra, Nucl. Phys. B **331**, 421 (1990);
A. E. Faraggi, Nucl. Phys. B **428**, 111 (1994) [hep-ph/9403312]; Phys. Lett. B **398**, 88 (1997) [hep-ph/9611219];
J. R. Ellis, A. E. Faraggi and D. V. Nanopoulos, Phys. Lett. B **419**, 123 (1998) [hep-th/9709049].

[25] L. M. Krauss and F. Wilczek, Phys. Rev. Lett. **62**, 1221 (1989).

[26] L. E. Ibanez and G. G. Ross, Phys. Lett. B **260** (1991) 291; Nucl. Phys. B **368**, 3 (1992);
T. Banks and M. Dine, Phys. Rev. D **45**, 1424 (1992) [hep-th/9109045];
C. Csaki and H. Murayama, Nucl. Phys. B **515**, 114 (1998) [hep-th/9710105].

[27] S. Kachru and E. Silverstein, Nucl. Phys. B **504**, 272 (1997) [hep-th/9704185].